REMARKS/ARGUMENTS

These remarks are submitted responsive to the office action dated March 31, 2004 (Office Action). A one month extension of time is herein requested along with the appropriate associated extension fee.

In paragraph 1 of the Office Action, the drawings were objected to under 37 C.F.R. § 1.83(a) because the drawings must show every feature of the invention as specified in the claims. Specifically, the Examiner asserted that "the graphical format and words or characters must be shown."

In response, Applicants have modified FIG. 3 to illustrate the claimed limitations. An attached replacement sheet and annotated sheet showing these changes to FIG. 3 have been included herein. Further, Applicants have amended the paragraphs from page 11, line 29 to page 12, line 13 as expressed in the amendments to specification section to correspond to the amended drawings. These modifications and amendments are supported by the claims and by the details contained within the Applications' specification. Accordingly, no new matter has been added. Applicants believe this amendment to the drawings and specifications corrects the deficiencies.

In paragraphs 2, the Examiner objected to claims 1, 7, 12, and 18 due to the lack of antecedent basis for "said user" and "said candidates." The phrases in the claims have been amended for proper antecedent basis.

In paragraphs 3-4 of the Office Action, the Examiner has rejected claims 1, 2, 6, 11, 12, 13, 17, and 22 under 35 U.S.C. § 102(e) as being anticipated by U.S. Patent No. 6,571,210 to Hon, et al. (Hon). In paragraphs 5-6, the Examiner has rejected claims 1-22 under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 5,842,163 to Weintrab (Weintrab) in view of U.S. Patent No. 5,712,957 to Waibel, et al. (Waibel).

Prior to addressing the rejections on the art, a brief review of the Applicants' invention is in order. The subject matter of the present invention includes a method for performing speech recognition. In particular, the invention can determine that a high likelihood exists that a recognition result does not accurately reflect received user speech. When such a determination is made, alternative word candidates can be automatically presented to the user for selection.

U.S. Appln. No. 09/858,399
Amendment Dated Aug. 2 2004
Reply to Office Action of Mar. 31, 2004
Docket No. 6169-234

IBM Docket No. BOC9-2000-0090

When the determination indicates that there is a high likelihood that the word is correctly identified (or not a high likelihood of a misrecognition), no presentation of alternative word candidates will occur. That is, an "accuracy threshold" can be used to determine whether an alternative word list is to be presented or whether the candidate having the topmost likelihood score is to be automatically accepted.

The presentation of alternative candidates can be limited to those candidates having a confidence score over a designated threshold. Accordingly, when the accuracy threshold determination indicates that an alternative list of candidates is to be presented, the presented candidates will be based upon one or more subsequent threshold determination(s), which compare the alternative candidates with a "presentation" threshold.

In one embodiment, instead of presenting the alternative word candidates, a user can be queried to identify one of the possible word candidates based upon a candidate identifier. The identifier need not be a complete expression of the candidate, but can instead be a identifier used to differentiate one potential candidate from another. For example, if the recognition results show there is an uncertainty as to whether the first letter of a spoken word was an "F" or an "S", the query (which can be an audible query from a voice response system) can ask "was that an F as in Frank or an S as in Sam," to which the user can respond. One of the potential word candidates can then be selected over other potential candidates based upon the query responses.

Turning to the rejections on the art, the Examiner has rejected claims 1, 2, 6, 11, 12, 13, 17, and 22 under 35 U.S.C. § 102(e) as being anticipated by Hon. Hon discloses a method and system of calculating a confidence measure based upon statistical comparisons between an N-best list of potential word candidates and a series of near-miss confidence templates. One such statistical comparison is shown at column 11, line 15. Hon teaches that comparisons are to be based upon multiple ratios and/or values, as shown in FIG. 8A-C and clearly stated throughout Hon. That is, Hon attempts to increase recognition accuracy and reject "noise" or out-of-vocabulary (OOV) utterances using confidence ratio(s) based on multi-element comparisons, like comparing the set of elements in the n-best list with the set of elements in one or more near-miss lists. As noted at column 2, lines 1-6, Hon is designed to overcome the short comings with

previous models that are based upon single ratio comparisons. Hon overcomes these limitations by statistically comparing multiple elements between two or more list of elements.

Referring to claims 1 and 12, Applicants claim the step of:

automatically presenting selected ones of said plurality of candidates as **alternative interpretations** of said speech **if none** of said confidence scores **is greater than said predetermined threshold.**

Referring to claims 11 and 22, Applicants claim the step of:

if said conference score is less than said minimum threshold, **presenting** at least one **word candidate as an alternative interpretation of said speech**, said word candidate being determined by a speech recognition engine **based upon said user speech and a confidence score**.

Hon fails to teach the step of conditionally (based upon confidence score comparisons) **presenting** candidates as alternative interpretations of said speech. Instead, Hon **never** teaches or suggests **presenting alternative candidates** (to a user). Instead Hon teaches that comparisons among multi-element templates (not single threshold values). Further, Hon never teaches that **alternative candidates** are to be **conditionally presented** to a user.

The Examiner cites column 12, lines 30-34 of Hon, which teaches that "a selected (*based upon an programmatically established parameter)* number of the normalized values and associated entries (herein, five) are retained to for the near-miss pattern 170, while all others are compiled into the value 174G at step 228 *(sic. Hon incorrectly states step 128)*." Then, the near-miss pattern is compared with the near-miss confident template (step 230) to accept or reject a hypothesis word based on comparison (step 30 of FIG. 1). As written, Hon cannot present alternative candidates since the step of accepting or rejecting a hypothesis word based on comparisons (step 30) occurs after step 228 of FIG. 10. Appreciably, alternatives (to a hypothesis word) **cannot** be presented until after the hypothesis word is accepted or rejected.

Further, Hon teaches away from the Applicants invention, which is to conditionally prompt a user to select from potential candidates when there is a high likelihood of a misrecognition. Hon is narrowly directed towards a particular statistical algorithm only applicable when performing multi-ratio comparisons between multiple elements in multiple word

{WP186528;1}   10

lists. Since Hon is clearly directed towards improving aspects of speech recognition (in particular perceived flaws in previous comparison algorithms) that are unrelated to and conflicting with the present invention, one of ordinary skill in the art would not turn to Hon for teachings pertaining to the present invention. That is, even though Hon is in the field of speech processing, the teachings of Hon are within a non-analogous subfield from that of the present invention.

Referring to claims 2 and 13, Applicants teach "selecting one of the candidates (to be presented to the user) that have confidence scores above a predetermined minimum threshold. Hon fails to teach or suggest this limitation. In fact, Hon teaches away from single element comparisons (like comparing an alternative candidates' confidence score against the minimum threshold) altogether. Hon does not teach a means for selecting candidates (to present) based upon such single element comparisons.

In light of the above, the 35 U.S.C. § 102(e) rejections of claims 1, 2, 6, 11, 12, 13, 17, and 22 should be withdrawn, which action is respectfully requested.

In paragraphs 5-6, the Examiner has rejected claims 1-22 under 35 U.S.C. § 103(a) as being unpatentable over Weintrab in view of Waibel. Weintrab discloses a method for determining the likelihood of appearance of keywords in a spoken utterance as part of a keyword spotting system of a speech recognizer, whereby a scoring technique is provided wherein a confidence score is computed as a probability of observing the keyword in a sequence of words given the observations. The technique involves hypothesizing a keyword whenever it appears in any of a plurality of "N-best" word lists with a confidence score that is computed by summing the likelihoods for all hypotheses that contain the keyword, normalized by dividing by the sum of all hypothesis likelihoods in the "N-best" list. In this regard, Weintraub like Hon, provides a speech processing technique where a multi-element list is compared against a plurality of other multi-element lists, which contradicts the teachings of the Applicants, claimed herein. Further, Weintraub, like Hon, **never** teaches or suggests conditionally (based upon single ratio comparisons) **presenting alternative candidates** to a user, as noted in paragraph 6, (lines 4-6) of the Office Action.

Waibel discloses a method and apparatus for repairing speech recognized by a recognition engine. Waibel generates at least two different N-best lists. The two N-bests lists are combined into a third N-best list. Errors in previously generated lists are "repaired" using replacements found in the third N-best list. Waibel, like Weintraub and Hon, **never** teaches or suggests conditionally **presenting alternative candidates** to a user.

In the Office Action, the Examiner cites FIG. 2 of Waibel as teaching "presenting selected ones of the plurality of candidates to a user as alternative interpretations of speech if none of the confidence scores is greater than a predetermined threshold. Waibel provides no such teachings. As noted at column 5, lines 6-26 of Waibel, FIG. 2 is a "flow chart" that shows the repair paradigm. It is a tabular example showing an information flow through which an utterance is repaired by Waibel's speech processing engine. The repair paradigm or "flow" is an information flow that occurs within repair module 12. No presentation of information in any form is taught, suggested, or contemplated by FIG. 2.

Other portions of Waibel, however, clarify that the only time a user is presented with a speech-recognized items is for purposes of confirming that a top selection was appropriate. The confirmation is not related to comparing potential candidates to thresholds in any fashion. The confirmation never presents a plurality of alternative candidates to a user. As shown in FIG. 3 of Waibel, only one substring (the most likely one selected by the speech engine) is to be "accepted by user 51." This is a traditional confirmation prompt, with a re-computation of a speech-recognition event occurring should the confirmation be negative. Similarly, FIG. 4 (item 70) and FIG. 5 (Item 70) shows that only one substring representing a top choice is conveyed to a user for confirmation.

Consequently, neither Weintraub, Waibel, nor any combination thereof teach or suggest presenting a **plurality of candidates** as **alternative interpretations** of said speech. Further, neither Weintraub, Waibel, nor any combination thereof teach conditionally taking any action (including presenting candidates to users) based upon results of comparing confidence scores within an N-best list against a single value (a predetermined threshold).

Moreover, Weintraub and Waibel both provide methods for selecting among a multitude of different N-best lists, which the Applicants do not teach. Instead, Applicants teach comparing entries in a single N-best list (performing single value comparisons) against a predetermined threshold, which Weintraub and Waibel fail to teach.

Referring to claims 1, 7, 11, 12, 18, and 22, Weintraub and Waibel are silent in regard to presenting alternative candidates to a user. Each of these claims, explicitly includes limitations for conditionally **presenting alternative candidates** to a user (via an interface) as an alternative interpretation of a "top choice" of a speech recognition operation. The condition of the conditional presentation is based upon a comparison of confidence values in an N-best list against a predetermined threshold.

Referring to claims 3, 5, 14, and 16, column 7, lines 6-18 and FIG. 4 of Waibel is cited to "show that a user input is received to indicate a user selection of a correct candidate from a plurality of **presented alternative candidates**." Waibel, however, does not present a plurality of candidates to a user. Instead, as shown from column 6, line 31 to column 7, line 15, automated routines select a best choice, where the automated routines occur in the repair module 12 and the speech recognition engine 14. The user is prompted to confirm this "best choice" in item 70. No alternative to the "most current best choice" of item 70 is presented to the user according to the teachings of Waibel.

Referring to claims 4 and 15, the Examiner took Official Notice that candidates are to be presented in a graphical format. Official Notice of this fact in light of the teachings of Weintraub and Waibel is inappropriate. Neither Weintraub or Waibel teach or suggest **selecting** particular ones of alternative candidates. Further, neither Weintraub or Waibel teach **presenting** alternative candidates to a user, such as presenting candidates in a graphical format.

Referring to claims 5 and 16, the Examiner failed to respond to the claimed limitation specifying that candidates are to be presented in an audio user interface, which is not taught or suggested by Weintraub, Waibel, nor any combination thereof.

In light of the above, the 35 U.S.C. § 103(a) rejections of claims 1-22 should be withdrawn, which action is respectfully requested.

The Applicants believe that this application is now in full condition for allowance, which action is respectfully requested. The Applicants request that the Examiner call the undersigned if clarification is needed on any matter within this Amendment, or if the Examiner believes a telephone interview would expedite the prosecution of the subject application to completion.

Respectfully submitted,

Date: 2 August 2004

Gregory A. Nelson, Registration No. 30,577
Kevin T. Cuenot, Registration No. 46,283
Brian K. Buchheit, Registration No. 52,667
AKERMAN SENTERFITT
Post Office Box 3188
West Palm Beach, FL 33402-3188
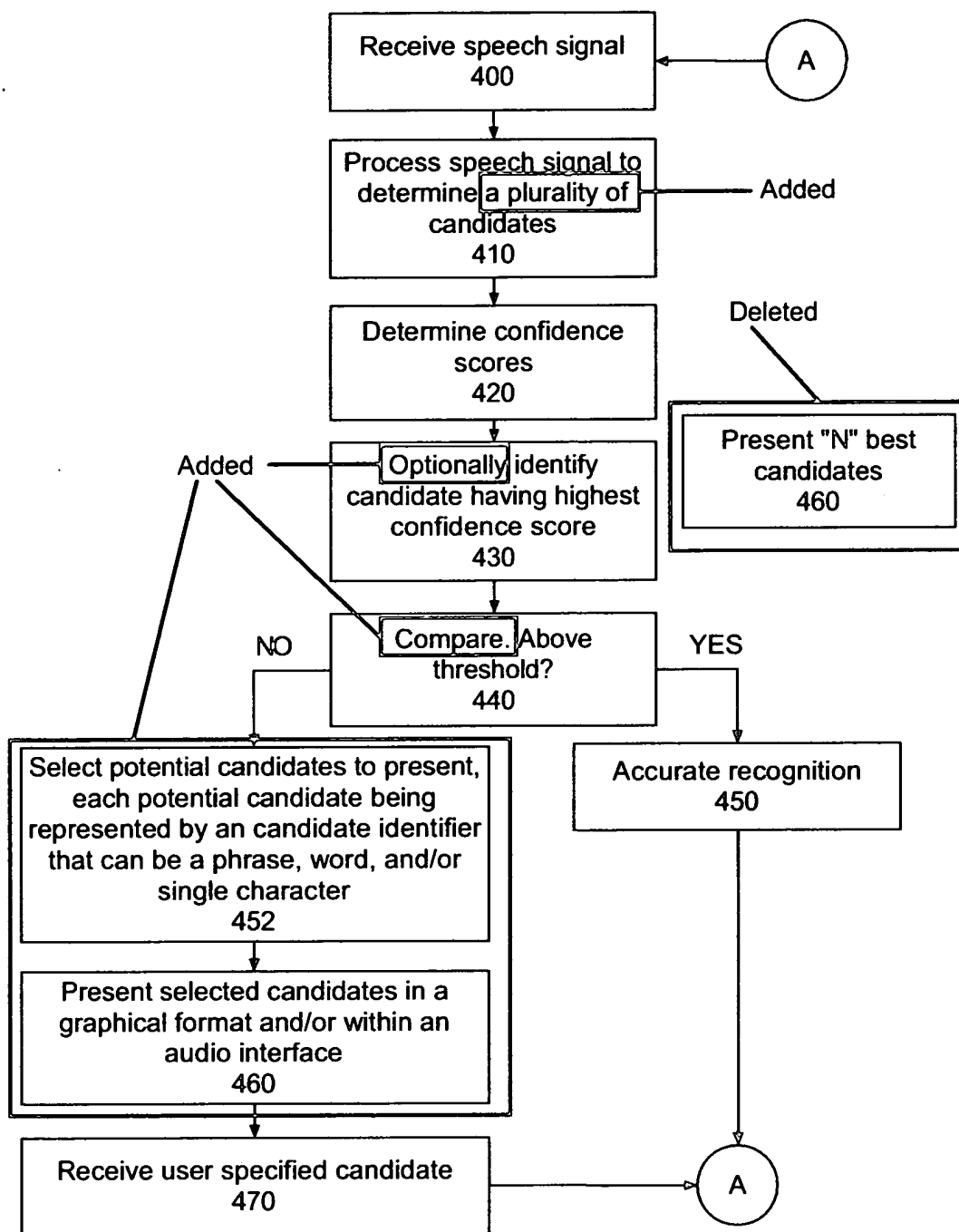Telephone: (561) 653-5000

Receive speech signal
400

A

Process speech signal to
determine a plurality of
candidates
410

— Added

Determine confidence
scores
420

Deleted

Present "N" best
candidates
460

Added —

Optionally identify
candidate having highest
confidence score
430

NO

Compare. Above
threshold?
440

YES

Select potential candidates to present,
each potential candidate being
represented by an candidate identifier
that can be a phrase, word, and/or
single character
452

Accurate recognition
450

Present selected candidates in a
graphical format and/or within an
audio interface
460

Receive user specified candidate
470

A

FIG. 3